Cassandra Crossing/ IA: fabbriche di informazioni false

(535)—Per descrivere correttamente i pericoli delle false Intelligenze Artificiali servono descrizioni tecniche, ma semplificate il più…

Cassandra Crossing/ IA: fabbriche di informazioni false

Figure 1:

(535)—Per descrivere correttamente i pericoli delle false Intelligenze Artificiali servono descrizioni tecniche, ma semplificate il più possibile.

3 aprile 2023— D'accordo, avrete ormai ben capito che Cassandra teme oggi le false "Intelligenze Artificiali" come, in un lontano passato, ha temuto i Greci e coloro che recano doni.

Purtroppo il dibattito che imperversa in ogni dove sulle false "Intelligenze Artificiali" , dai salotti televisivi alle pubblicazioni universitarie, impedisce la comprensione, e talvolta addirittura la percezione, del loro pericolo e della sua enormità.

Tentiamo quindi di semplificare al massimo il discorso tecnico per farne emergere la sostanza, che è per fortuna molto semplice. Per favore, se siete arrivati fino qui, leggetemi fino in fondo.

Tanti, ma davvero tanti anni fa, per descrivere gli sforzi degli informatici di far progredire i computer di allora, fu coniato il fortunatissimo ed oggi devastante termine di "Intelligenza Artificiale".

Un campo di ricerca che doveva migliorare la comprensione delle domande che gli umani ponevano ai computer, di permettere a questi ultimi di capire e rispondere in linguaggio naturale, e di generalizzare i problemi al fine di essere in grado di rispondere a qualsiasi domanda.

Un tale obbiettivo, molto remoto e probabilmente irraggiungibile, veniva identificato con il termine "Intelligenza Artificiale Generale", per distinguerlo dai risultati, molto parziali e deludenti, che decenni di ricerca su questo tema avevano fornito.

Poi sono nate le reti neurali, e dopo molto tempo i supercomputer. Dall'unione di questi due progressi, sono nate le tecnologie chiamate "*Deep Learning*"—Apprendimento Profondo, altro termine fuorviante, ma lasciamo perdere per semplicità.

Queste tecnologie permettono, una volta che gli siano stati fornite grandi moli di dati, di rilevarne automaticamente certe caratteristiche e di estrarne alcune informazioni. Usando queste informazioni estratte, sono in grado di produrre una risposta se le vengono presentate ulteriori informazioni.

Un esempio, molto abusato, è che se gli vengono fornite grandi quantità di foto di gattini e di altre cose, è in grado di calcolare la "***gattità***" o meno di una foto, e di dire se una nuova foto è di un gatto oppure no con un'alta probabilità di azzeccarci. Questo ha permesso di realizzare applicazioni che funzionano bene, ma solo se se usate in campi ristrettissimi, ad esempio per classificare come tumorali o meno radiografie ed immagini di RMN o TC.

Successivamente sono state create applicazioni in campo grafico e linguistico, in grado di generare non semplicemente un si/no, ma una "risposta" grafica o testuale, ad una domanda posta in linguaggio naturale.

Sono sempre software che necessitano di grandi quantità di dati e di potenza di calcolo per essere "allenati", e vengono chiamati "modelli di linguaggio generativi".

Questi software deducono molto bene le regole grammaticali, sintattiche ed altre caratteristiche dei vari linguaggi umani e perfino dei linguaggi per computer, e sono quindi in grado di "proseguire" una frase che gli venga fornita in ingresso, creando un discorso di grande verosimiglianza.

Se la frase iniziale è una domanda, forniscono una sequenza di parole, anzi un discorso molto ben formato, che ha spesso tutte le caratteristiche di una risposta credibile.

Ed è vero, **tutte le caratteristiche di una risposta, tranne una, l'esattezza**. La vogliamo chiamare "Verità"? La vogliamo chiamare "Conoscenza"?

E veniamo finalmente all'attualità.

Un software prodotto dall'azienda OpenAI, chiamato **GPT-3 "Generative Pre-trained Transformer"**—Trasformatore generativo pre-allenato, dopo essere stato nutrito con una quantità enorme di dati, si è rivelato particolarmente efficace nel generare risposte credibili, ed è stato messo a disposizione del pubblico tramite un'interfaccia colloquiale chiamata **ChatGPT**.

ChatGPT ha avuto un'enorme risonanza, e moltissime persone, dopo averla provata, hanno cominciato a scrivere cose indicibili, minchiate galattiche, descrizioni tanto fantasiose quanto errate su questo software, sui suoi successori e sulle tecnologia alla loro base, descrivendola come il futuro paradiso dell'umanità, come la soluzione a tutti i problemi.

Come oracoli di verità.

Se questo articolo di Cassandra fosse un romanzo giallo, a questo punto il lettore attento potrebbe già individuare l'assassino. Ma continuiamo nella storia, ed arriviamo al finale.

Cassandra, come profetessa, ha frequentato gli oracoli, e ne ha sperimentato tutti i limiti. Un famoso responso fornito da un oracolo, la Sibilla romana, ad un soldato in procinto di partire per la guerra, che voleva sapere se ne sarebbe tornato, fu "***Ibis, redibis non morieris in bello.***"

Per chi non avesse avuto frequentazioni con la lingua latina, la sequenza di sei parole, mancando di una virgola, poteva significare sia quello che il soldato voleva sentirsi dire "*Partirai, tornerai, non morirai in guerra*", sia il suo opposto "*Partirai, non tornerai, morirai in guerra*". Non conteneva nessuna informazione. Era una "non-risposta".

Un risposta priva di significato e di informazione, ma a cui il soldato fornì il significato che desiderava, cioè che sarebbe sopravvissuto, e non quello che sarebbe morto in guerra.

Ma qualcuno potrebbe pensare che l'oracolo conoscesse davvero il futuro, e si divertisse malignamente a renderlo comunque oscuro ai poveri mortali.

Così certamente non è per ChatGPT, perché questo software non solo non conosce il futuro, ma nemmeno conosce la risposta a nessuna domanda, nemmeno se una foto sia di un gatto oppure no.

ChatGPT e tutti i suoi successori dello stesso tipo non conoscono, non conoscono per niente, non conoscono assolutamente, non conoscono neanche lontanamente, non conoscono per niente la risposta a qualsiasi domanda; non conoscono assolutamente nulla. La conoscenza con cui vengono "nutriti" viene "digerita" e svanisce, lasciando solo un credibile "**Mentecatto Artificiale**" privo di qualsiasi Intelligenza o Cultura.

ChatGPT e tutti i suoi successori sono fatti solo ed unicamente per generare risposte "credibili" a domande poste in linguaggio naturale.

Risposte credibili, non risposte vere.

Questo concetto deve essere recitato come un mantra, ripetuto fino allo sfinimento, pubblicizzato in ogni occasione.

ChatGPT e tutti i suoi successori non hanno nessuna conoscenza sulla verità o falsità di una domanda, non hanno nessuna cultura, pur avendone digerito quantità enormi. Sono un vicolo cieco, **utilizzabile solo come arma di propaganda di massa. E lo stanno facendo!**

Sono solo in grado di darvi risposte credibili, in cui voi individuerete il significato che volete o potete dargli.

E qui sta il pericolo, questi oggetti vengono propagandati come risponditori di domande, non come affabulatori o generatori di prosa fantasiosa.

E la loro attuale linea di sviluppo non è quella di renderli capaci di risposte esatte, cosa che non sarà mai possibile, ma solo incapaci di dare risposte così errate da essere evidenti o ridicole, risposte che siano espresse in termini razzisti, sessisti, in violazione di qualsiasi possibile concetto "politically correct".

Non c'è quindi nessuna possibilità, se non per puro caso, di avere risposte informate e corrette da modelli linguistici, cioè da software come GPT-3, ChatGPT, GPT-4 ed i loro fratelli e successori.

Non sono fatti per questo. Non lo sanno e non lo possono fare per progetto.

Non conoscono, e non sanno calcolare, categorie come "verità" o "esattezza"

E' assolutamente impossibile che forniscano risposte esatte ed affidabili, per quanti "miglioramenti" e filtri vi vengano aggiunti

Nel frattempo la maggioranza del pubblico, **indotta in errore da un'attività di propaganda commerciale che non si è mai vista nel mondo delle tecnologie informatiche, è convinta di interagire con un software che può fornire vere risposte**, mentre in realtà è in grado di fornire solo sequenze di parole ben formate, ma assolutamente vuote di qualsiasi contenuto affidabile. Parole messe a caso ma con grande proprietà di linguaggio.

Potrei dire che possono produrre solo discorsi da politicanti e venditori di auto usate, ma fate conto che non l'abbia detto.

Speriamo che questa ennesima ripetizione, spero ben argomentata ripetizione, serva a qualcosa.

Se vi affidate a ChatGPT, non tornerete mai dalla guerra.

Scrivere a Cassandra—Twitter—Mastodon
Videorubrica "Quattro chiacchiere con Cassandra"
Lo Slog (Static Blog) di Cassandra
L'archivio di Cassandra: scuola, formazione e pensiero

By Marco A. L. Calamari on April 3, 2023.

Canonical link

Exported from Medium on August 27, 2025.